AWARD NUMBER: **W81XWH-14-1-0374**

TITLE: **Autism and Obesity: Co-Occurring Conditions or Drug Side Effects?**

PRINCIPAL INVESTIGATOR: **Zohreh Talebizadeh, PhD**

CONTRACTING ORGANIZATION: **Children's Mercy Hospital**
Kansas City, MO 64108

REPORT DATE: **October 2015**

TYPE OF REPORT: **Annual**

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE						*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE October 2015	**2. REPORT TYPE** Annual	**3. DATES COVERED** 30 Sep 2014 - 29 Sep 2015

4. TITLE AND SUBTITLE Autism and Obesity: Co-Occurring Conditions or Drug Side Effects?	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER **W81XWH-14-1-0374**
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Zohreh Talebizadeh, PhD E-Mail: ztalebi@cmh.edu	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Children's' Mercy Hospital Kansas City, MO 64108-4698	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

[1] IRB protocol has been submitted and an Exempt status has been determined for the project. [2] Weight/Height data has been obtained from the AGRE database and BMI has been calculated for autistic subjects (n=552). Autistic subjects were then classified based on the BMI data into obese, overweight, healthy and underweight categories. [3] Using a rigorous literature search dbSNP search, we collected data on 157 AIWG SNPs and used multiple tools and selection criteria to identify Tag SNPs for these SNPs. [4] We found SNAP (SNP Annotation and Proxy Search) tool to be the most useful for selecting Tag SNPs. [5] We downloaded AGRE genotyping data (combined Affymetrix and Illumina platforms). We are examining this genotyping list for the presence of the AIWG Master List SNPs and the identified Tag SNPs. [6] We began developing a statistical pipeline for analyzing the genotyping data (bioequivalence tests). A poster was presented at the annual ASHG meeting summarizing the current status of our statistical method.

15. SUBJECT TERMS

AGRE, AIWG, ASD, BMI, SSC, SNP, KCALSI, PCORI, ASHG

16. SECURITY CLASSIFICATION OF: **U**			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USAMRMC
a. REPORT U Unclassified	**b. ABSTRACT** U Unclassified	**c. THIS PAGE** U Unclassified	UU Unclassified	11	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

1. **INTRODUCTION:**

The **objective** of this pilot project is to better understand the relationship between autism and obesity. It is not clear if obesity is co-occurring with autism or is related to antipsychotic-induced weight gain (AIWG). Weight gain is one of the main side-effects of the commonly used antipsychotics. Since the majority of patients with autism take antipsychotics, a *general assumption* is that the observed elevated rate of obesity in autism (i.e., 40%) is caused by AIWG. Our **hypothesis** is that the prevalence of known AIWG associated SNPs in obese and non-obese autistic subjects is comparable; thus, AIWG cannot be the only reason for the observed higher rate of obesity. To test this hypothesis, we will re-analyze already existing data (from AGRE and SSC families) by comparing the prevalence of AIWG associated SNPs in obese and non-obese autistic subjects.

2. **KEYWORDS:**

AGRE: Autism Genetic Resource Exchange
AIWG: Antipsychotic-Induced Weight Gain
ASD: Autism Spectrum Disorder
BMI: Body Mass Index
SSC: Simons Simplex Collection
SNP: Single Nucleotide Polymorphism

ASHG: American Society of Human Genetics
KCALSI: Kansas City Area Life Science Institute
PCORI: Patient-Centered Outcomes Research Institute

3. **ACCOMPLISHMENTS:**

- **What were the major goals of the project?**

Task 1. Identification of autistic subjects, Month 1-18
Percentage of Completion: 65%

Task 2. Finding known AIWG SNPs from existing genetic datasets, Month 1-18
Percentage of Completion: 40%

Task 3. Identification of Tag SNPs, Month 1-18
Percentage of Completion: 40%

Task 4. Evaluating the AIWG SNPs genotyping profiles in the discovery cohort (n=200), Month 13-24
Percentage of Completion: 20%

Task 5. Replicating statistical findings in a second independent validation cohort (n=800), Month 13-24
Percentage of Completion: Nothing to Report

Task 6. Finalizing analyses and preparing reports and manuscripts, Month 18-24
Percentage of Completion: 30%

- **What was accomplished under these goals?**

Task 1.

- IRB protocol was submitted and an Exempt Status has been issued for this study.
- Protocol for access to the SSC database has been prepared / submitted (**Status:** under review).
- Weight/Height data has been obtained from the AGRE database and BMI has been calculated for autistic subjects (the PI is an approved AGRE investigator) (See summary Table below):

Table-BMI distribution-AGRE autistic subjects

BMI categories	TOTAL (%)
Obese (≥95th)	126 (23%)
Overweight (85th to <95th)	83 (15%)
Healthy (5th to <85th)	300 (54%)
Underweight (<5th)	43 (8%)
ALL	**552**

Task 2.

1. We have been able to compile data on 157 AIWG SNPs, which cover all the reported SNPs, so far, and we refer to this list as our Master List in this report.
2. We correlated the SNPs in our Master List with the dbSNP database, a public-domain archive for genetic polymorphisms to collect more detail information, including physical mapping, population data, and microarray platforms. Here is a brief summary of dbSNP information for the 157 SNPs (Master List):
    - 153 were found in dbSNP
    - 146 included in Illumina or Affymetrix microarray platforms
    - 11 SNPS are not handled by Illumina or Affymetrix
    - 2 SNPs did not map to any genome assembly
    - 1 SNP has an invalid snp_id value

Task 3.

We have already begun identification of Tag SNPs using multiple resources (**Status:** ongoing)

1. We utilized several tools to obtain Tag SNPs: SNAP, TagSNP, and Tagger. Following is a brief summary of Tag SNPs information for the 157 AIWG SNPs (Master List):

2. We uploaded our SNPs Master list to the Tag SNP tool and found 98 Tag SNPs.

3. We found SNAP (SNP Annotation and Proxy Search) tool to be the most useful for selecting Tag SNPs. There are several criteria included in SNAP to specify/narrow down a search.

For example,

Search criteria (#1):

- SNP data set: HapMap3 (release2)
- Population panel: CEU
- r2 threshold: 0.8
- Distance limit: 500
- Include each query snp as a proxy for itself

- Select all arrays
- Apply array filter to: query SNPs and proxy SNPs,

Search Result (#1):

- *84 Tag SNPs found for 84/157 Master list SNPs*

Search criteria (#2):

- SNP data set: 1000 Genomes Pilot 1
- Population panel: CEU
- r2 threshold: 0.8
- Distance limit: 500
- Include each query snp as a proxy for itself
- Select all arrays
- Apply array filter to: query SNPs and proxy SNPs,

Search Result (#2):

- *862 Proxy SNPs (203 duplicates, 731 unique values) for 129 Master list SNPs were found,*
- *28 SNPs were not found*

Search criteria (#3):

- SNP data set: 1000 Genomes Pilot 1
- Population panel: CEU
- r2 threshold: no limits
- Distance limit: 500
- Include each query snp as a proxy for itself
- Select all arrays
- Apply array filter to: query SNPs and proxy SNPs,

Search Result (#3):

- *1463 Proxy SNPs (1369 duplicates, 94 unique values) found for 109 Master list SNPs*

We also searched for LD status of our Master list SNPs (Method: Pairwise LD Search tool in SNAP).
Search criteria (LD):

- SNP data set: 1000 Genomes Pilot 1
- Population panel: CEU
- r2 threshold: 0.8
- Distance limit: 500
- Include each query snp as a proxy for itself
- Select all arrays
- Apply array filter to: query SNPs and proxy SNPs,

Search Result (LD):

- *36 SNPs Proxy SNPs found for 28 Master list SNPs*

4. For the SNPs from Master list (n=28), for which Tag SNPs were not found by SNAP in 1000GenomesPilot1, we run them individually through the dbSNP database.

Search Result:

- 610 neighbor SNPs (53 duplicates, 557 unique values) found for 22/28 SNPs
- neighbor SNPs not found for 6/28 SNPs

5. Additionally, our consultant, Dr. Mueller, provided us with a list of 23 high/low priority AIWG SNPs that is being considered in his studies. Six of them are not among our Master list SNPs because the papers that report those SNPs are still in press. We are making a separate list for these high/low priority AIWG SNPs and plan to identify Tag SNPs following the methods applied on our Master List.

Task 4.

- We downloaded AGRE genotyping data (combined Affymetrix and Illumina platforms)
    - N= 16303 SNPs
    - **Status:** we are examining this genotyping list for the presence of the AIWG Master List SNPs and identified Tag SNPs

- We began developing a statistical pipeline for analyzing the genotyping data (bioequivalence tests) (**Status:** ongoing). Here is a brief summary of the statistical modeling:

Over the past year, the co-investigator and Senior Biostatistician, Daisy Dai, has been working diligently with PI, Zohreh Talebizadeh, on development of appropriate statistical methods for this project. Our aim is to prove potential bioequivalence for SNPs that have been considered as risk factors for AIWG in obese vs non-obese autistic cohorts. To do so, we have conducted extensive literature review for bioequivalence tests in genetic studies. We found very limited methods in this field. In contrast, there is rich literature with a large body of publications for methods to identify risk factors and genetic association. Therefore, we have been focusing on comparing existing methods and evaluate their advantages and disadvantages using empirical assessment. Due to the complexity in genetic data, we need to create multiple scenarios by taking varying correlation structures and effect sizes (no effect, very small effect, moderate effect) into account. The empirical assessment will randomly generate 10,000 data sets. For each data set, multiple statistical methods will be assessed. We will calculate Type I error rate and power for each scenario.

Task 5.

Nothing to Report

Task 6.

- We presented a poster at the ASHG meeting describing the statistical pipeline we are developing for conducting bioequivalence tests on genotyping (SNP) data:

  Zohreh Talebizadeh, Hongying Dai, Ayten Shah. Equivalence tests for the analysis of genotyping data: Assessing equality of SNPs in study cohorts (Abstract/Program #1437). Presented at the 65th Annual Meeting of the American Society of Human Genetics (ASHG), October 8, 2015, Location: Baltimore, MD (see Appendix- PDF copy of the poster)

- **What opportunities for training and professional development has the project provided?**

"Nothing to Report"

- **How were the results disseminated to communities of interest?**

"Nothing to Report"

- **What do you plan to do during the next reporting period to accomplish the goals?**
  At this point we do not anticipate a major change in our approach and SOW. Therefore, for the next reporting period we are going to continue with the remaining tasks as described in our original SOW.

4. **IMPACT:**

- **What was the impact on the development of the principal discipline(s) of the project?**

"Nothing to Report"

- **What was the impact on other disciplines?**

Using the conceptual strategy developed in this DOD project (i.e., reanalyzing existing genetic data to address an important question related to a patient population (i.e., is obesity a drug side effect or co-morbidity in autism?), the PI was able to design and submit a grant application to PCORI. The Engagement aspect of this new research plan has been funded, and we are submitting a full Methods application for implementation of our unique conceptual strategy to improve outcomes research projects, which we plan to apply it on three conditions (autism, cancer, and cardiovascular diseases).

- **What was the impact on technology transfer?**

"Nothing to Report"

- **What was the impact on society beyond science and technology?**

"Nothing to Report"

5. **CHANGES/PROBLEMS:**

"Nothing to Report" Please note the explanation provided below for "Actual or anticipated delays"

- **Changes in approach and reasons for change**
- **Actual or anticipated problems or delays and actions or plans to resolve them**
  The PI has been administratively moved under a different Division in our Institution since September 2015. This administrative change did not impact her research performance and resources; however, there has been an unexpected delay in processing of some pending invoices, including consultant fees and meeting expenses (ASHG), which will be resolved soon.

- **Changes that had a significant impact on expenditures**
- **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**
- **Significant changes in use or care of human subjects**
- **Significant changes in use or care of vertebrate animals.**
- **Significant changes in use of biohazards and/or select agents**

6. **PRODUCTS:**

"Nothing to Report"

- **Publications, conference papers, and presentations**
  "Nothing to Report"

- **Journal publications.**

"Nothing to Report"

- **Books or other non-periodical, one-time publications.**
  "Nothing to Report"

- **Other publications, conference papers, and presentations.**

Zohreh Talebizadeh, Hongying Dai, Ayten Shah. Equivalence tests for the analysis of genotyping data: Assessing equality of SNPs in study cohorts (Abstract/Program #1437). Presented at the 65th Annual Meeting of the American Society of Human Genetics (ASHG), October 8, 2015, Location: Baltimore, MD (see Appendix- PDF copy of the poster)

- **Website(s) or other Internet site(s)**
  "Nothing to Report"

- **Technologies or techniques**
  "Nothing to Report"

- **Inventions, patent applications, and/or licenses**
  "Nothing to Report"

- **Other Products**
  "Nothing to Report"

7. **PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS**

- **What individuals have worked on the project?**

Name:	Zohreh Talebizadeh "no change"
Name:	Ayten Shah "no change"
Name:	Daisy Dai "no change"

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

Here are updates for Zohreh Talebizadeh (PI):

Patton Trust Research Development Grants-KCALSI 2013-2014 (Role: PI)
Title: "*Analysis of circadian genes in autism: characterization of alternative splicing profile of JARID1 (KDM5) genes*". **Status:** Ended

Two New Active Projects (no overlap with the DOD project):

1. University of Missouri System (**$100.000**) 2015-2016 (Role: **co-investigator**)
   Title: "*Epigenetic and immune factors in the effect of maternal stress exposure on autism*"
   The goal of this study is to study gene-environment interaction in autism, focusing on maternal stress exposure.

2. Patient-Centered Outcomes Research Institute (PCORI)-Engagement Award (EAIN-2419) **($50,000)** 2015-2016 **(**Role: **PI)**
   Title: "*Incorporating genetic data in PCOR studies: building a road map for stakeholder engagement*"

The goal of this project is to leverage communication between a wide range of stakeholders with diverse backgrounds to assess IF/How known genetic risk factors can be incorporated in patients' outcomes research projects.

- **What other organizations were involved as partners?**

"Nothing to Report"

8. **SPECIAL REPORTING REQUIREMENTS**

- **COLLABORATIVE AWARDS:**

Not Applicable

- **QUAD CHARTS:**

Not Applicable

9. **APPENDICES:** *Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc. Reminder: Pages shall be consecutively numbered throughout the report.* ***DO NOT RENUMBER PAGES IN THE APPENDICES.***

Simulating the dataset involves three steps: (1) modeling genotype data, (2) modeling disease risks, and (3) modeling disease status.

UMKC School of Medicine

Equivalence tests for the analysis of genotyping data: Assessing equality of SNPs in study cohorts

Children's Mercy HOSPITALS & CLINICS www.childrensmercy.org

Zohreh Talebizadeh, Hongying Dai & Ayten Shah

Children's Mercy Hospital & University of Missouri-Kansas City School of Medicine, Kansas City, MO

Abstract

Statistical testing is strictly based on null (Ho) and alternative (Ha) hypotheses. The construction of statistical hypotheses will determine the interpretation of results. In most analyses of genotyping data or genome-wide association studies, the null hypotheses assume there is no association between SNP and phenotype. Rejection of null hypotheses will provide strong evidence to indicate the potential associations between SNP and phenotype. The aim of our study is to demonstrate the application of equivalence tests on genotyping data in situations that require testing the presence of equality instead of differences. A series of SNP data will be generated and tested for a hypothetical scenario; i.e., rule out the impact of tested SNPs on a given comorbidity in a disease group, by demonstrating equality between corresponding means. The constructed Ho is: there is association between SNP and comorbidity in the patient population versus the Ha: there is no association between SNP and comorbidity in the patient population. An equivalence test is warranted to test the constructed hypotheses.

Equivalence test has been widely applied in clinical trials to confirm the equivalency in drug efficacy (i.e., bioequivalency), but rarely on genotype data. In this work we will discuss: 1) differences between equivalence test and differential test, 2) misuse of differential test for equivalence testing in the context of analyzing genotyping data, 3) minimal sample size to establish an equivalence limit. The impact of variables such as: allele frequency, sample size, and the number of tested SNPs on equivalence of cases and controls will also be assessed.

We will perform extensive simulation study to illustrate that misuse of differential tests may cause bias in stating equality of SNPs between cohorts. For genotyping data and genome-wide association analysis, a limited number of equivalence tests are available as compared to very rich pool of differential tests.

Funding Agency. Department of Defense (AR130398)

Introduction

Chen et al., 2000 proposed tests for equivalence or non-inferiority between two proportions. Their methods were originally designed to evaluate bioequivalence between two treatments or two drugs by comparing the success rates or eradication rates of binomial outcome variables. Their methods have not been applied in genotype data.

For our genotype data, we plan to compare the minor allele (%), minor genotype type (%). This method cannot be directly applied to assess bioequivalence for haplotypes or diplotypes. We will consider whether we can extend the methods to these two areas. The test can be performed by testing two sets of one-sided hypothesis each at the nominal level α (Type I error rate). To conclude equivalence, both hypotheses need to be rejected. Consequently, the two one-sided tests form a test with an overall type I error rate α. Alternatively, it is generally equivalent to comparing the confidence limits on the difference of the two means with the equivalence limits π.

The objective of the test-reference comparison is to demonstrate the "similarity" of minor allele/genotype between two groups (1 & 2). In the bioequivalence test, the null hypothesis is that the difference in minor allele/genotype (%) is no smaller than a predetermined limit. For instance, if we set the predetermined limit of minor allele difference is 5% for SNP rsXXX. Then the null hypothesis is that the difference in minor allele (%) of rsXXX is ≥ 5%. If in group 1, the minor allele is 2% and in group 2, the minor allele is 3%. Then we can reject the null hypothesis and claim bioequivalence of rsXXX between groups 1 and 2.

Conclusion

Our study will contribute to addressing this gap by providing a useful protocol, including examples, for application of such tests on genotyping data. More equivalence tests need to be developed to fulfill the needs of genotype testing analysis.

Methods (simulation of genotyping data)

Simulating the dataset involves three steps: (1) modeling genotype data, (2) modeling disease risks, and (3) modeling disease status. We simulated genotyping data using R, commonly used statistical software (https://www.r-project.org). GenABEL, or *ABEL, is an umbrella name for a number of software packages aiming to facilitate statistical analyses of polymorphic genome data. It is a rich program set which now allows very flexible genome-wide association (GWA) analysis (GenABEL, ProbABEL, MixABEL, OmicABEL), meta-analysis (MetABEL), parallelization of GWA analyses (ParallABEL), management of very large files (DatABEL), and facilitates evaluation of prediction (PredictABEL).

We then used "simulatedDataset" in the "PredictABEL" package to simulate the genotyping data. In the "simulatedDataset" function, we defined the following parameters:

- **ORfreq: Matrix with ORs and frequencies of the genetic variants. The matrix contains four columns in which the first two describe ORs and the last two describe the corresponding frequencies. The number of rows in this matrix is same as the number of genetic variants included. Genetic variants can be specified as per genotype, per allele, or as dominant/ recessive effect of the risk allele. When per genotype data are used, OR of the heterozygous and homozygous risk genotypes are mentioned in the first two columns and the corresponding genotype frequencies are mentioned in the last two columns. When per allele data are used, the OR and frequency of the risk allele are specified in the first and third column and the remaining two cells are coded as '1'. Similarly, when dominant/ recessive effects of the risk alleles are used, the OR and frequency of the dominant/ recessive variant are specified in the first and third column, and the remaining two cells are coded as '0'.**
- **Poprisk: population disease risk (expressed in proportion).**
- **Popsize: total number of individuals included in the dataset.**

The simulation method assumes that (i) the combined effect of the genetic variants on disease risk follows a multiplicative (log additive) risk model; (ii) genetic variants inherit independently, that is no linkage disequilibrium between the variants; (iii) genetic variants have independent effects on the disease risk, which indicates no interaction among variants; and (iv) all genotypes and allele proportions are in Hardy-Weinberg equilibrium. Assumption (ii) and (iv) are used to generate the genotype data, and assumption (ii) and (iii) are used to calculate disease risk.

Brief method description

Reference

Chen, J.J., et al (2000). Tests for eqivalence or noninferiority between two proportions. Drug Information Journal *34*, 569-578.

Figure 1. Data simulation framework

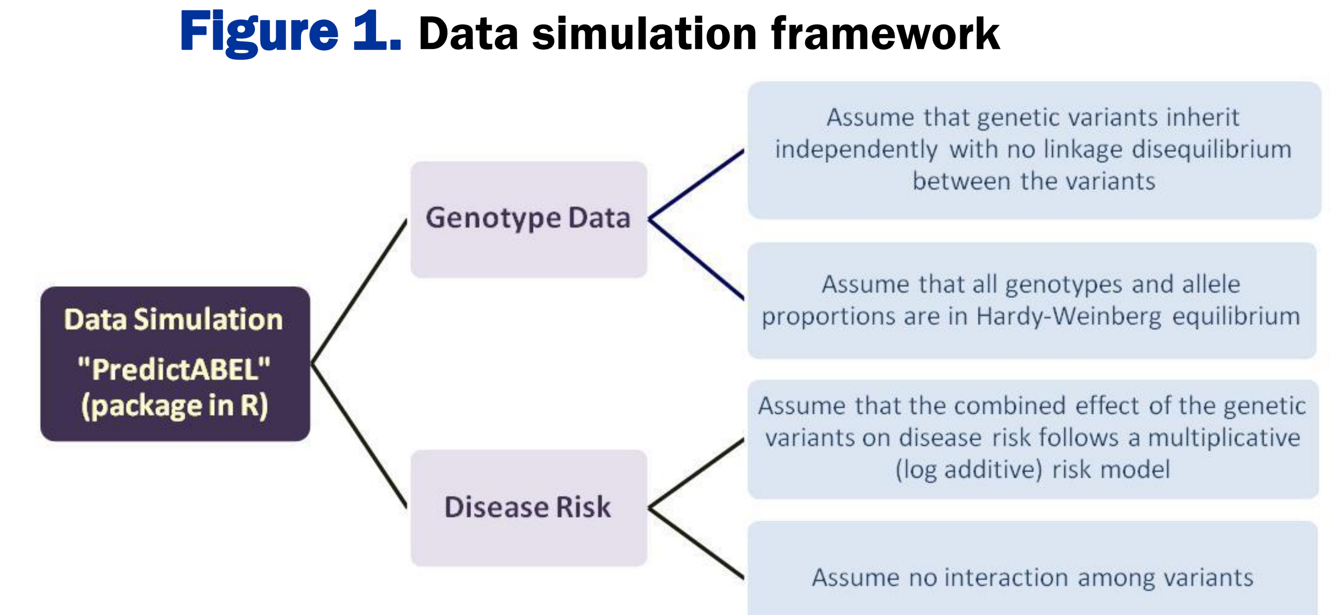
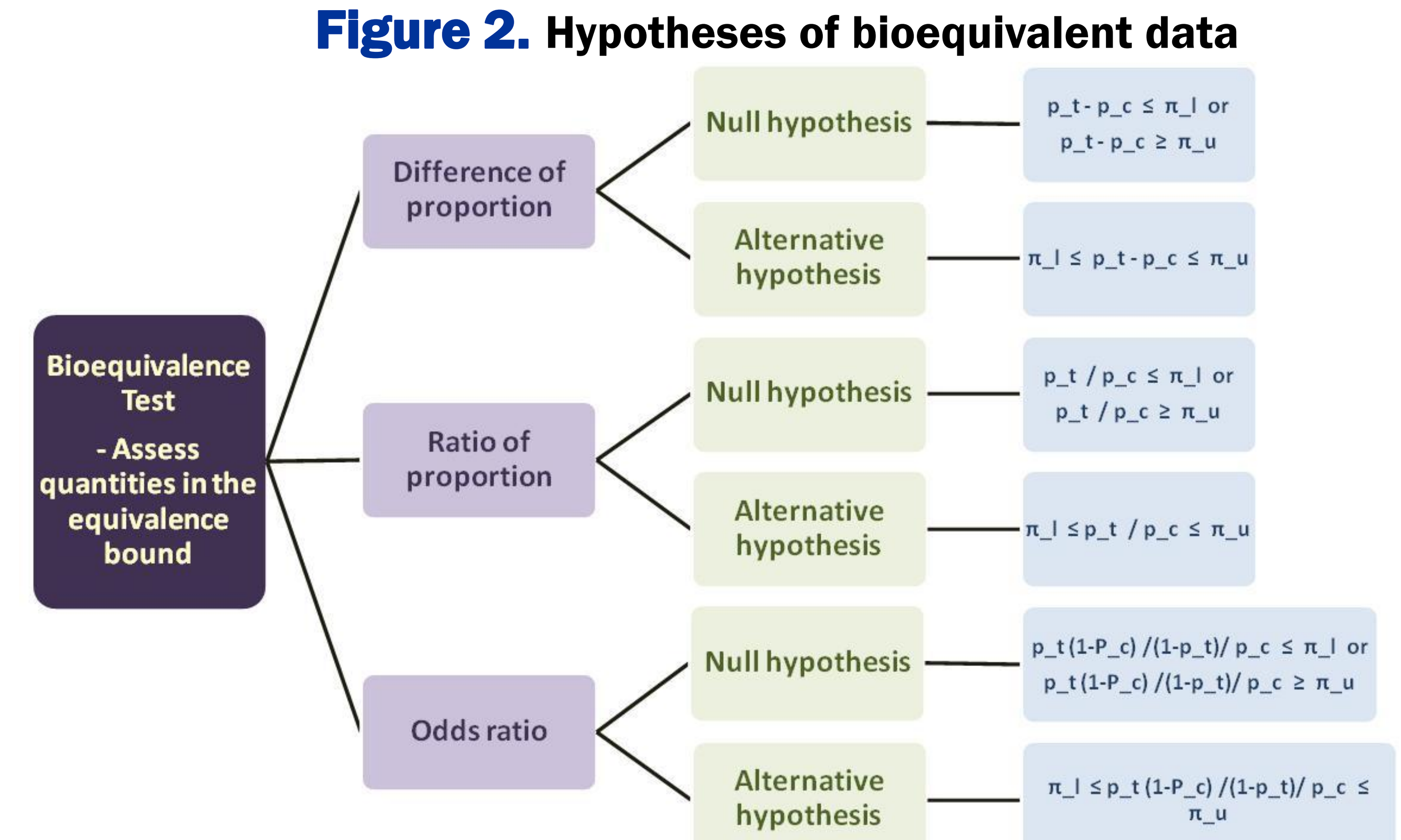


Figure 2. Hypotheses of bioequivalent data



AIM: To investigate bioequivalence test for general genotyping data

• **We have simulated genotyping data using "PredictABEL".**

• **The simulation of genotype data method assumes that genetic variants inherit independently with no linkage disequilibrium between the variants; and all genotypes and allele proportions are in Hardy-Weinberg equilibrium.**

• **The simulation of disease risk data assumes that the combined effect of the genetic variants on disease risk follows a multiplicative (log additive) risk model and that genetic variants have independent effects on the disease risk, which indicates no interaction among variants.**

Figure 3. Rejection region